

University of Dundee

## Argumentation in the 2016 US Presidential Elections

Visser, Jacobus; Konat, Barbara; Duthie, Rory; Koszowy, Marcin ; Budzynska, Katarzyna;  
Reed, Chris

*Published in:*  
Language Resources and Evaluation

*DOI:*  
[10.1007/s10579-019-09446-8](https://doi.org/10.1007/s10579-019-09446-8)

*Publication date:*  
2019

*Licence:*  
CC BY

*Document Version*  
Publisher's PDF, also known as Version of record

[Link to publication in Discovery Research Portal](#)

### *Citation for published version (APA):*

Visser, J., Konat, B., Duthie, R., Koszowy, M., Budzynska, K., & Reed, C. (2019). Argumentation in the 2016 US Presidential Elections: Annotated corpora of television debates and social media reaction. *Language Resources and Evaluation*, 54, 123-154. <https://doi.org/10.1007/s10579-019-09446-8>

### General rights

Copyright and moral rights for the publications made accessible in Discovery Research Portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from Discovery Research Portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain.
- You may freely distribute the URL identifying the publication in the public portal.

### Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



# Argumentation in the 2016 US presidential elections: annotated corpora of television debates and social media reaction

Jacky Visser<sup>1</sup> · Barbara Konat<sup>2</sup> ·  
Rory Duthie<sup>1</sup> · Marcin Koszowy<sup>2,3</sup> ·  
Katarzyna Budzynska<sup>1,2</sup> · Chris Reed<sup>1</sup>

© The Author(s) 2019

**Abstract** In this paper we present US2016, the largest publicly available set of corpora of annotated dialogical argumentation. The annotation covers argumentative relations, dialogue acts and pragmatic features. The corpora comprise transcriptions of television debates leading up to the 2016 US presidential elections, and reactions to the debates on Reddit. These two constitutive parts of the corpora are integrated by means of the intertextual correspondence between them. The rhetorical richness and high argument density of the communicative context results in cross-genre corpora that are robust resources for the study of the dialogical dynamics of argumentation in three ways: first, in empirical strands of research in discourse analysis and argumentation studies; second, in the burgeoning field of argument mining where automatic techniques require such data; and third, in formulating algorithmic techniques for sensemaking through the development of Argument Analytics.

**Keywords** Argumentation · Corpus · Intertextual correspondence · Political discourse · Reddit · Television debate · US elections

---

This research was supported in part by the Engineering and Physical Sciences Research Council in the UK under Grants EP/M506497/1 and EP/N014871/1, and in part by the Polish National Science Centre under Grant 2015/18/M/HS1/00620.

---

✉ Jacky Visser  
j.visser@dundee.ac.uk

<sup>1</sup> Centre for Argument Technology, University of Dundee, Dundee DD1 4HN, UK

<sup>2</sup> Institute of Philosophy and Sociology, Polish Academy of Sciences, Nowy Swiat 72, 00-330 Warsaw, Poland

<sup>3</sup> University of Białystok, Mickiewicza 1, 15-213 Białystok, Poland

# 1 Introduction

Argument and debate are as ubiquitous as they are fundamental to the functioning of society. In philosophy, the theory of argumentation has been studied as a distinct field since the 1960s (though its heritage can be traced back much farther), whilst in linguistics and computational linguistics, it has only become a focus much more recently. One of the key challenges facing empirically driven research in argumentation (both in more theoretically oriented strains of linguistics as well as practically driven research in natural language processing) is the need for appropriate data and, typically, annotated data. The lack of data has been severely hampering such research and has been hobbling development in the nascent field of argument mining in particular. The dearth of such resources is rooted in two key challenges: first, the technical challenge of distilling the rich work of argumentation theory into a theoretically coherent approach which can be translated into a practical set of annotation guidelines; and second, the prosaic challenge of the labour-intensive nature of annotation, particularly given that it typically requires training and is not, in general, delegable to crowdsourced solutions.

In this paper, we describe the largest publicly available corpus of argumentatively annotated debate which makes use of a detailed approach to argument analysis founded upon an integration of the leading philosophical approaches to dialogical argument (Reed and Budzynska 2011). The data comprises transcripts of televised political debates leading up to the 2016 presidential election in the United States of America: *viz.*, the first Republican primary debate, the first Democrat primary debate, and the first general election debate between Hillary Clinton and Donald Trump. In addition, we include precisely contemporaneous reaction online, and in particular, from the social media platform Reddit. This lays the scene for an unusually rich dataset, which not only captures dialogical interaction (as opposed to monological—and often artificially generated—argument which is much more common), but also allows exploration of reaction in social media. This connection offers the opportunity for the first time to investigate cross-genre and intertextual connections using empirically robust methodology, and also allows exploration of the relationships between, on the one hand, topics, structures and arrangements of arguments, and, on the other, their reception with a larger audience.

We proceed by first introducing the domain of discourse in more detail (Sect. 2), to subsequently describe the data selection and annotation methodology (Sects. 3 and 4)—including how the resulting corpus can be accessed (Sect. 4.2), and how the annotation has been validated (Sect. 4.3)—and explore the notion of ‘intertextuality’ (as introduced in Sect. 5.1) and the benefits to be gained by connecting the annotated transcripts of live television debates with associated social media reactions (Sect. 5). Finally, brief indication is offered of the types of research benefiting from the newly developed resource (Sect. 6), and how this relates to the existing literature (Sect. 7).

## 2 Argumentation in television debates and social media

### 2.1 Argumentation in discourse

The corpus that we present in this paper deals with argumentation in political communication. Argumentation refers to the appeal to reasoning in discourse in support of a contested point of view (van Eemeren et al. 2014). If two interlocutors find themselves in disagreement about the acceptability of a standpoint, arguments can be used to resolve this disagreement in a reasonable way by testing the reasons supporting it. The standpoint at issue could be an opinion, a belief, a proposal, or anything else the interlocutors might disagree about that could be resolved through a reasoned exchange of arguments and criticisms. The reasons, or arguments, put forward as part of such a discussion can be structured in various ways and can draw on a broad range of inferential reasoning principles.

As a case in point, consider the argumentative defence by Marco Rubio of his standpoint that he should be the Republican party's nominee for the 2016 presidential elections in the US. As part of a neatly structured series of statements about economic changes and the need for forward-looking candidates, who understand the actual problems of citizens, such as living paycheck to paycheck and having student debts, Rubio maintains that the focus of the Republican party in the elections should be on the future, not on the past. After alluding to the sufficiency of his own resume, Rubio makes the case in Example (1) that focusing on past achievements would be detrimental to the electoral chances of a Republican candidate.<sup>1</sup> In (1), Rubio employs the explicit discourse marker “because” to signal his supporting argument that the resume of Hillary Clinton, whom he presumes will be the Democratic candidate, is better than that of any of the Republican hopefuls (van Eemeren et al. 2007; Das and Taboada 2017).

- (1) Marco Rubio: [...] *if this election is a resume competition, then Hillary Clinton's gonna be the next president, because she's been in office and in government longer than anybody else running here tonight.*

In Fig. 1, the simple structure of Rubio's argumentation is visualised as a diagram (in a graph-based format that will be explained in more detail in Sect. 3.1). The diagram shows the propositions that are the content of Rubio's utterances and the inferential relations between them.<sup>2</sup> If his audience accepts the bottom proposition—the premise—as well as the reasoning principle underpinning the

<sup>1</sup> Example (1)—taken from our annotation of the first Republican primaries television debate on 6 August 2015 in Cleveland, Ohio—is available online at [www.aifdb.org/argview/10828](http://www.aifdb.org/argview/10828).

<sup>2</sup> The inferential relation is reflected in the diagram by means of a *Default Inference* node. ‘Default inference’ indicates that there is some form of argumentative support or justification happening, while the specific inferential principle that the argumentation relies on is not specified further: such annotation of ‘argument(ation) schemes’ (Walton et al. 2008; van Eemeren et al. 2014) is not the object of our current exposé—but it is explored elsewhere by Visser et al. (2018b). The further introduction in Sect. 3 will show that the same holds for the transitions and conflict and rephrase relations: all of these are currently represented in their default forms.

argumentative inference (which in this instance is in reverse temporal order), then this constitutes a successful defence of the top proposition—the conclusion—which subsequently contributes to his audience accepting Rubio’s standpoint, i.e. that he is the right candidate for the future.

Our characterisation of argumentation implies that both its propositional dimension of the underlying logical reasoning structure, and its dialogical dimension of the linguistic realisation in communication should be taken into account. Returning to Example (1), Fig. 2 shows how the reasoning appealed to in Rubio’s argumentation is anchored in the structure of the dialogue.<sup>3</sup> The dialogical context of Rubio first asserting the controversial “if this election is a resume competition, then Hillary Clinton’s gonna be the next president” to be followed by “she’s been in office and in government longer than anybody else running here tonight”, is what makes the latter assertion into an argument in defence of the first assertion (explicitly signalled with “because”). With a different dialogical embedding, Rubio’s locutions (his contributions to the dialogue) could play different communicative roles, for example as an explanation or as part of a question-answer sequence. Our conception of argumentation (see Sect. 3.1) allows for the representation of both the propositional and the dialogical dimensions (as seen respectively on the left side and the right side of Fig. 2), integrated by means of the communicative functions of the locutions (see Sect. 3.1), such as the *Asserting* and *Arguing* in Fig. 2.

## 2.2 Argumentation in televised election debates

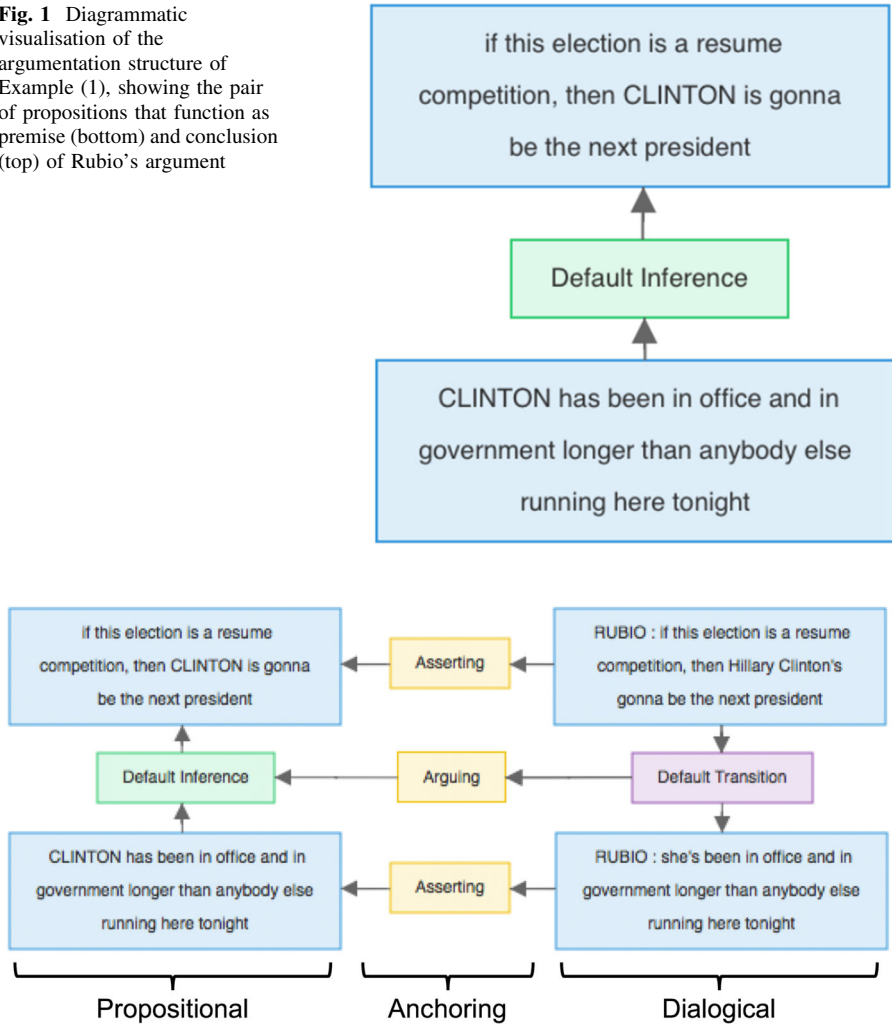
The US2016 corpus comprises transcripts of televised debates for the 2016 presidential elections in the United States of America. Ever since the first televised election debate between the then US presidential candidates John F. Kennedy and Richard Nixon in 1960, the debates have played an important role in the democratic process in many countries (Kraus 2013). The general election and the corresponding television debates between Hillary Clinton and Donald Trump as the candidates from the two dominant political parties in the US (respectively the Democratic Party and the Republican Party) took place in the Autumn of 2016.<sup>4</sup> Prior to the general elections, both main parties held primary elections and caucuses to elect their presidential candidate. These primaries were also preceded by television debates between the leading prospective candidates in 2015 and 2016.

While the format of each of the debates is slightly different, there are some recurring characteristics. The television networks’ moderators pose questions to the invited candidates, and guide the debate (for example by keeping time and order), while the candidates make opening statements, answer the moderators’ (and occasionally the public’s) questions, defend their views and challenge those of their political opponents, in an attempt to garner more support among the electorate. For

<sup>3</sup> The dialogical annotation of Example (1) is available online in OVA (see Sect. 3.3) at [www.arg.tech/ova-10828](http://www.arg.tech/ova-10828).

<sup>4</sup> In the current paper, we focus exclusively on the debates between the (prospective) candidates of the two dominant parties in US politics.

**Fig. 1** Diagrammatic visualisation of the argumentation structure of Example (1), showing the pair of propositions that function as premise (bottom) and conclusion (top) of Rubio's argument



**Fig. 2** Diagrammatic visualisation of the analysis of Example (1), showing how the propositional reasoning on the left is anchored in the dialogical realisation of the argument on the right

the general elections, three television debates were organised between Democratic candidate Clinton and Republican candidate Trump, and one debate between their candidate vice-presidents. For the primaries, the Republican party held 12 debates for the front-runners and seven so-called 'undercard' debates between the next tier of candidates. The Democratic party held 10 primary debates.

The context of televised election debates fosters a mixture of well-structured and well-presented argumentation that appears to have been prepared in advance, and impromptu argumentation originating from the need to cope with the interactional dynamics. The latter poses a challenge in the analysis of the argumentation.

Consider Example (2), advanced by then prospective candidate (now President) Trump.<sup>5</sup> Trump anticipates his claim about the topic of immigration to not be accepted outright. He therefore supports it with multiple statements. Upon closer inspection, Trump's support relies mostly on the rhetorical device of repetition, with several of his assertions constituting a relation of rephrase rather than inference. By relying on varying ways of presenting the same content within a superficially inferential reasoning structure, Trump introduces an element of circularity.

- (2) Donald Trump: *So, if it weren't for me, you wouldn't even be talking about illegal immigration, Chris. You wouldn't even be talking about it. This was not a subject that was on anybody's mind until I brought it up at my announcement. And I said, Mexico is sending. Except the reporters, because they're a very dishonest lot, generally speaking, in the world of politics, they didn't cover my statement the way I said it.*

### 2.3 Argumentation in political social media discussions

In addition to the transcripts of televised election debates, the US2016 corpus contains annotated social media reactions to these debates. In particular, we look at the responses on the social media platform Reddit. This 'second screen' interaction moves the audience from a passive role as consumer into an active role as participant in a multi-genre conversation across communicative mediums. Not only does this serve as a predictor for political involvement (Gil de Zuniga et al. 2015), live interactions with televised material can actually be a means for increasing citizen's engagement (Plüss and De Liddo 2015). The items on social media are related to the television debates not only through the topics that are addressed (i.e. either the topics that are being discussed in the television debate, or what happens in the debate as a topic itself), but also by the time at which the online discussion takes place (i.e. live reaction while the television debate is going on).

Reddit is an online discussion platform ([www.reddit.com](http://www.reddit.com)) with between 10 million and 18 million unique users per month. The user community is organised to areas of interest, called 'subreddits', dedicated to a great variety of topics, ranging from the discussion of the aesthetic qualities of celebrities to technological issues, and from culinary advice to politics. The messages in the subreddits are organised in threads, comprising a tree structure, with threads containing a large number of comments being referred to as 'megathreads'.

The written online discourse on Reddit can be contributed to by anyone who is a registered user of the social media platform (as long as they do not violate the user agreement). This means that a greater diversity in language use is to be expected (within the boundaries of the explicit etiquette guidelines), with people contributing from varying backgrounds, nationalities and education levels. The different

<sup>5</sup> Example (2)—also taken from our annotation of the first Republican primaries television debate on 6 August 2015 in Cleveland, Ohio—is available online at [www.aifdb.org/argview/10829](http://www.aifdb.org/argview/10829).

vocabulary and style is, for example, evident in the frequent use of profanities (both in usernames and in posted comments).

These considerations lead to an expectation of a mixed argumentative quality (in terms of both rhetorical efficacy, and of dialectical and logical fallaciousness) in the online discussions, with potentially many less well-crafted and well-signalled examples. Further complicating the annotation is the lack of a moderator that enforces turn-taking. This means that contributions to the online dialogue can come in rapid succession, with posts sometimes responding simultaneously to the same previous item, complicating the interpretation of referents. Despite the potential difficulties (and common grammatical and typing mistakes), it is clear that the Reddit discussions also contain clearly argumentative, well-structured content. In (3), for example, Reddit user Bigtwinkie supports an evaluative standpoint about the uncomfortably hostile nature of the first television debate for the Democratic primaries, by drawing an analogy to a domestic scenario that is relatable to the audience.<sup>6</sup>

- (3) Bigtwinkie: *This debate has honestly been making me uncomfortable, its been way too hostile. Its like listening to mom and dad fight in the kitchen while your hide under the covers in your room.*

### 3 Methodology

#### 3.1 Theoretical foundations

The annotation of the corpus is based on Inference Anchoring Theory (IAT) (Budzynska and Reed 2011; Reed and Budzynska 2011). Building on insights from Discourse Analysis and Argumentation Theory, IAT offers an explanation of argumentative conduct in terms of the anchoring of reasoning structures in persuasive dialogical interactions: bridging the *logical* reasoning dimension, and the *dialogical* communicative dimension of argumentation. In the summarised IAT annotation guidelines (Sect. 3.2), we provide further details on the key terminology introduced in the explanation of the theoretical backgrounds of IAT.

The reasoning appealed to in the argumentation involves three types of relations between propositions. First, an *inference relation* holds between a proposition that is meant to function as a premise in an argument and the contested proposition that it supports as a conclusion. Second, a *conflict relation* indicates that one proposition is understood to be incompatible with another. Third, a *rephrase relation* is intended to hold between propositions that are similar (in both content and argumentative function) but not identical. Although these relations could in principle exist between propositions regardless of dialogical embedding—e.g. two propositions *p* and *not-p*

<sup>6</sup> Example (3)—taken from our annotation of the Reddit reaction to the first Democratic primaries television debate on 13 October 2015 in Las Vegas, Nevada—is available online at [www.aifdb.org/argview/10058](http://www.aifdb.org/argview/10058).



contradicting, or  $p$  being entailed by  $q$ , or  $p'$  being a paraphrase of  $p$ —in IAT such relations are only considered relevant if anchored in discourse. This means that each of the inference, conflict and rephrase relations is actualised in discourse by the interlocutors' utterances. These utterances are conceived of as a sequence of *locutions* by one or more speakers, linked together by *transitions* reflecting the protocol that structures the dialogue (i.e., which locutions of a specific type are uttered at each particular stage of a dialogue; viz. 'adjacency pairs' (Sacks et al. 1974; Schegloff and Sacks 1974; Jacobs and Jackson 1982), and 'dialogue games' (e.g. Carlson 1983; Mann 1988; Walton and Krabbe 1995)).<sup>7</sup>

The propositions and relations that together form the argumentative reasoning are anchored in the locutions and transitions that constitute the dialogue by means of *illocutionary connections*. Elaborating on traditional Speech Act Theory (Austin 1962; Searle 1969), illocutionary forces are reinterpreted as relations connecting locutions (and transitions) to propositional content (and propositional relations). The illocutionary connection specifies the dialogical function that is *intended* to be applied to the propositional content: in other words, the act that is performed by means of the locution. For example, a speaker can *assert* that a proposition  $p$  is the case, or *question* whether  $p$  is true, and she can *argue* to invoke an inferential relation between two propositions  $p$  and  $q$  (functioning as a premise and a conclusion).

Distinctive of IAT is that it is a theory of argumentation geared towards computational linguistic methods and software implementation. To facilitate machine-readability, IAT adheres to the extended Argument Interchange Format (AIF<sup>+</sup>) standard (Chesñevar et al. 2006; Reed et al. 2008b). AIF<sup>+</sup> is a graph-based ontology that facilitates the representation of the intertwined locutionary, illocutionary, and propositional structures, resulting from the analysis of argumentative discourse. The ontology's information nodes (or I-nodes) are instantiated to represent propositions, and the locutions that are used to convey them. Various types of scheme nodes (or S-nodes) are employed to represent relations between I-nodes (and occasionally S-nodes): e.g., transitions between locutions, illocutionary connections between (for example) locutions and propositions, or inferences between propositions.

### 3.2 Summary of annotation guidelines

Four annotators were trained for 50 h at the Centre for Argument Technology at the University of Dundee in using IAT to analyse the television debates and Reddit discussions. They acquainted themselves with the communicative contexts of televised election debates and social media posts. Concurrently, they were taught about the foundations of IAT, practised individually with the use of the IAT-based guidelines for annotating discursive and argumentative structure on around 3000 words of election debate and Reddit discussion texts not part of the final corpus, and they discussed the resulting practice annotations amongst themselves and with

<sup>7</sup> We use the term 'speaker' for the producer of an utterance, whether spoken or written, and 'hearer' for the addressee.

expert annotators. The full annotation guidelines are available online at [www.arg.tech/US2016-guidelines](http://www.arg.tech/US2016-guidelines) and deal with, among other issues: anaphoric references, epistemic modalities, repetitions, punctuation, discourse indicators, interposed text, reported speech, and distinguishing between the pure, assertive, and rhetorical use of questions and challenges. Below, we provide a summary of the most important aspects of the annotation scheme.

The particular genre and communicative domain from which textual data is drawn inherently influence the annotation, and the annotators have to be sensitive to this influence. The television debates are annotated on the basis of a transcription of the spoken discourse. This means that some of the multi-modal features of the interaction are lost, making the text harder to interpret. The transcripts include questions from the general audience, sometimes introduced into the original debate by means of a video, posed to the candidates by the moderators. From the original Reddit posts as well, some content is lost. For example, if a user deletes his or her account, this does not delete any existing posts, but it does replace their username with ‘deleted’. Both the audience questions in the television debates and the deleted user accounts in the Reddit discussions result in unexpected speaker names showing up in the corpus: in the first case, the name of the audience member that poses the question will be added to the list of contributors to the debate; in the second case, *deleted* will occur as a participant—we should be aware however that this is not one individual user, but rather the collection of all users that deleted their account in the time between them posting the comment and us collecting the textual data.

In annotating the texts, the annotators followed an iterative procedure. The iterative nature is necessary to account for the interdependencies between the various analytical tasks. For example, the first consideration is manually segmenting the original text into locutions, but as the summarised guidelines below stipulate, the segmentation is partly based on argumentative functions, which requires further analysis of the text. This means that annotators go back and forth between different stages of the analysis to account for all the interdependencies. That being said, annotators will generally look to address the annotation tasks in the order in which we present the summarised guidelines below.

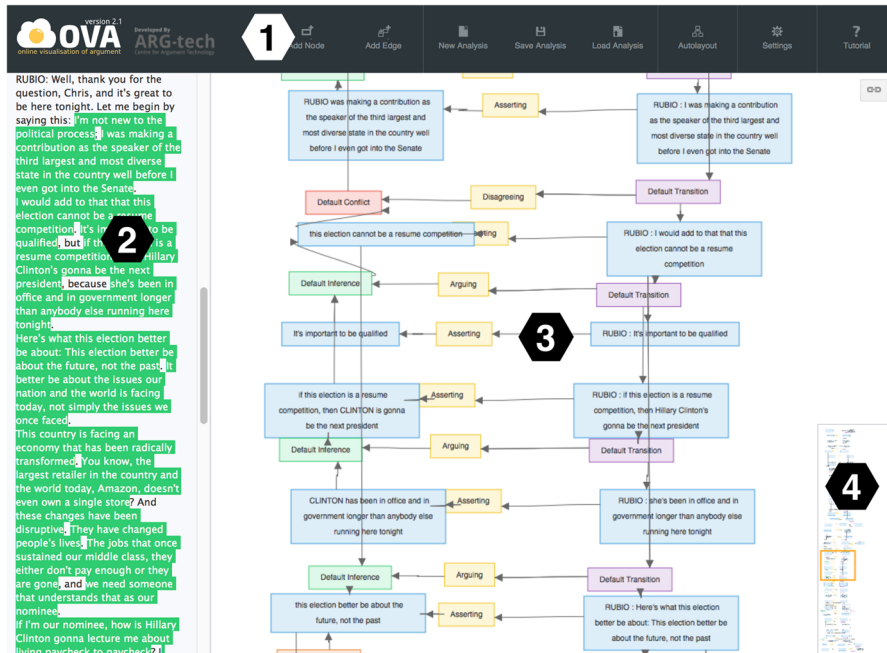
- *Locutions* A locution is the unit into which the (transcribed) text is segmented. A locution consists of a speaker designation and an argumentative discourse unit (ADU) (Peldszus and Stede 2013) in the following format: “SPEAKER : ADU”—see, e.g., the right top and bottom nodes in Fig. 2. An ADU is any text span that has a discrete argumentative function.<sup>8</sup>
- *Transitions* A transition captures the functional relationship between locutions—see, e.g., the right middle node in Fig. 2. The transitions reflect the protocol of the dialogue (or the structural rules of the ‘dialogue game’).

<sup>8</sup> ADUs are based on EDUs (‘elementary discourse units’), analytically relevant non-overlapping spans of text (although there are various interpretation of what exactly constitutes an EDU in the literature: Grimes (1975) and Givón (1983) view them as clauses while Hirschberg and Litman (1993) view them as prosodic units, Sacks et al. (1974) as conversational turns, Polanyi (1988) as sentences, and Grosz and Sidner (1986) as intentional discourse segments.)

- *Illocutions* An illocutionary connection embodies the intended communicative function of a locution or transition between locutions—see, e.g., the middle column of nodes in Fig. 2. Although other types do occur, most relevant types of illocutions within the context of US2016 are *Agreeing*, *Arguing*, *Asserting*, (*Pure/Assertive/Rhetorical*) *Challenging*, *Disagreeing*, (*Pure/Assertive/Rhetorical*) *Questioning*, *Restating*, and *Default Illocuting* (when none of the other types suffice; at present mainly in the case of question answering). Illocutionary connections anchor the propositional contents (see below) and relations between them (see below) in the locutions and transitions that constitute the discourse annotation.
- *Propositions* Depending on the type of illocutionary connection, a propositional content should be reconstructed—see, e.g., the left top and bottom nodes in Fig. 2.
- *Inferences* An inference is a directed relation between propositions, reflecting that a proposition is meant to supply a reason for accepting another proposition—see, e.g., the left middle node in Fig. 2. Such support may be annotated as instantiating a specific argument scheme (e.g., *Argument from Example* or *Argument from Expert Opinion*) or the annotated relation may default to *Default Inference*.
- *Conflicts* A conflict is a directed relation between propositions, reflecting that a proposition is meant to be incompatible with another proposition or propositional relation – the conflict would take up the same place as the *Default Inference* on the left middle in Fig. 2. Incompatibility between propositions may depend on, e.g., *Logical contradiction* or *Pragmatic contrariness*, or the annotated relation may default to *Default Conflict*.
- *Rephrases* A rephrase is a directed relation between propositions, reflecting that a proposition is meant to be a reformulation of another proposition—the rephrase would take up the same place as the *Default Inference* on the left middle in Fig. 2. Reformulation of propositions may involve, e.g., *Specialisation*, *Generalisation* or *Instantiation*, or the annotated relation may default to *Default Rephrase*.

### 3.3 Annotation software

In their work, the annotators made use of the OVA software (Janier et al. 2014), which is freely available at [www.oiva.arg.tech](http://www.oiva.arg.tech). OVA, or Online Visualisation of Argumentation, assists the analysis of argumentative discourse by allowing the user to visualise both the dialogical and the propositional structure of the argumentation within one software environment. It is well suited to support the annotation of discourse on the basis of IAT. Figure 3 shows a screen-shot of OVA: under the menu items (indicated with the label “1”) at the top, the left pane of the window shows the text transcript (label “2”) that is to be annotated (in this case from the first head-to-head between Clinton and Trump), the right pane is where the analytical structure goes (“3”), and the right bottom corner contains a navigation inset (“4”). As part of the segmentation of the text, annotators select a piece of text on the



**Fig. 3** Screen-shot of the OVA annotation software; the four numbered labels indicate, respectively, the menu bar [1], the original text [2], the annotation graph [3], and the navigation overview of the full argument map [4]

left, which gets added to the right as a node. By following the annotation guidelines, OVA then allows the node to be connected to others with new nodes and edges indicating, e.g., the discourse transitions between locutions, the illocutionary connections between locutions and propositions, and the propositional relations between propositions. The diagrams that result from an analysis with OVA are saved in the searchable online AIFdb repository of annotated argumentation ([www.aifdb.org](http://www.aifdb.org)), exploiting the AIF<sup>+</sup> compliance of IAT (Lawrence et al. 2012).

## 4 The US2016 corpus

### 4.1 Data collection

The transcripts of the television debates were collected from The American Presidency Project, a non-partisan online archive of over 124,000 documents related to the US presidency (Peters and Woolley 1999). To bring the overall amount of text down to a manageable level, we only took into consideration three of the (prospective) candidates' debates preceding the 2016 US presidential elections: the first of each series of debates for the primaries of the Republican and Democratic parties, and for the general elections. Our corpus contains annotated transcripts of

**Table 1** Overview of the composition of the US2016 corpus

<b>US2016</b>	<b>US2016R1</b>	<b>US2016D1</b>	<b>US2016G1</b>
<b>US2016tv</b>	US2016R1tv	US2016D1tv	US2016G1tv
<b>US2016reddit</b>	US2016R1reddit	US2016D1reddit	US2016G1reddit

the first Republican candidates debate for the primaries on 6 August 2015 in Cleveland, Ohio (Peters and Woolley 2015b), the first Democratic candidates debate for the primaries on 13 October 2015 in Las Vegas, Nevada (Peters and Woolley 2015a), and the first general election debate on 26 September 2016 in Hempstead, New York (Peters and Woolley 2016).

The Reddit material was manually retrieved from the Reddit website. To put some boundaries on the size of the relevant discourse on Reddit, we only took into account the mega-thread(s) that corresponded to the respective television debate while it took place. Every 30 min a new mega-thread was created on Reddit. From this abundance of discursive material, we selected sub-threads corresponding to specific time windows on the basis of the degree of dialogical interaction in the television debate.

For example, there was high dialogical interaction (expected to foster more argumentative online reaction) in the first general election television debate during the time window between 1:58:45 AM UTC and 2:05:45 AM UTC. We then selected sub-threads on Reddit which were posted between 1:58:45 AM UTC and 2:05:45 AM UTC. The thread and turn structures of the Reddit material were preserved while selecting the sub-threads that encompassed at least five dialogue turns. Pilot annotations showed that sub-threads shorter than five turns do generally not exhibit structured argumentative interaction. Because these short exchanges tend not to promote structured arguments or conflicts, we excluded them from our corpus. We also excluded sub-threads dedicated to jokes and wordplay. Due to the nature of the Reddit community, users sometimes post jokes and wordplays merely to gain attention or to elicit some emotional response. These posts mainly serve the phatic function of language, and typically lack argumentation or topical disagreement. Finally, we excluded technical threads and those not related to discussion of the television debates, such as those used to discuss technical or practical problems, either with Reddit itself or with the television broadcast.

## 4.2 Structure and availability of the corpus

The US2016 corpus comprises ‘argument maps’ as its constitutive units. An argument map is the result of the annotation of a conveniently sized excerpt of the analysed text, typically consisting of 500 to 1500 words. The argument maps that constitute the US2016 corpus are organised in several sub-corpora. Table 1 shows how the sub-corpora are compiled. The six corpora listed in boldface (US2016, US2016R1, US2016D1, US2016G1, US2016tv, and US2016reddit) in the top row and in the first column are derived from the other six corpora (US2016R1tv,

The screenshot shows the AIFdb Corpora interface. At the top, there is a search bar and the AIFdb logo. Below the header, the main section is titled "US2016". Underneath, there is a link to download the full corpus: [tar.gz](#) | [zip](#) || View [analytics](#). To the right of this link is a black hexagon with the number 3. Below this, there are two argument maps. The first is titled "Argument Map 10808" and contains the text: "not only will Americans be electing a new president next year Americans also will be electing a world leader voters should assess the candidate's experience, character and vision for the future voters make this important decision CHAFEE is the only one running for president that has been a mayor,a...". Below this map is a black hexagon with the number 1. To the right of the first map is a black hexagon with the number 2. The second argument map is titled "Argument Map 10809" and contains the text: "people are disgusted with the way that money has corrupted our political process, intimidating incumbents and empowering Wall Street every day, the turnstile government that we see, and also the power of the financial sector in both parties People are looking for a leader who understands how the...". Below this map is a black hexagon with the number 4. At the bottom, there is a third argument map titled "Argument Map 10810" with the text: "O'MALLEY has two minutes my name is Martin O'Malley, former Mayor of Baltimore, former governor of Maryland, a".

**Fig. 4** Screen-shot of the AIFdb Corpora interface; label [1] locates download functionality, [2] miniature argument diagrams, [3] access to Argument Analytics, [4] editing with OVA

US2016R1reddit, US2016D1tv, US2016D1reddit, US2016G1tv, and US2016G1reddit).

All corpora part of our 2016 US elections annotation project are identified with the 'US2016' prefix. The affix 'R', 'D', or 'G' indicates that the sub-corpus covers, respectively, the Republican primaries (R), the Democratic primaries (D), or the general election (G). The numbered affix indicates the position of the debate in the series of debates organised for the primaries, and for the general election. Because the corpus currently only contains texts relating to the first debates, all sub-corpora have the affix '1' (leaving open the possibility to extend the corpus at a later time). Finally, the suffix 'tv' or 'reddit' indicates whether the sub-corpus contains excerpts from the televised candidates' debates (tv) or from user contributed discussion on the Reddit social media platform (reddit). For example, the sub-corpus US2016R1reddit contains only and all of the transcripts in the corpus from the Reddit megathreads ('reddit') related to the first ('1') Republican primary debate ('R').

The derivative corpora are composed as follows. US2016 is the main corpus and contains all the other sub-corpora. US2016tv contains the sub-corpora of annotated televised debates, while US2016reddit contains the sub-corpora of annotated discussion on Reddit in relation to the televised debates. US2016R1 combines the sub-corpora of both the first televised Republican primaries debate and the corresponding discussion on Reddit; similarly US2016D1 and US2016G1 comprise

the cross-genre sub-corpora for, respectively, the first Democratic primaries debate, and the first general election debate.

The US2016 corpus, and all sub-corpora, are openly available online through AIFdb Corpora at [www.corpora.aifdb.org](http://www.corpora.aifdb.org) (Lawrence and Reed 2014). The (sub-) corpora can be directly accessed by adding the corpus' abbreviated name (see Table 1) to the AIFdb Corpora URL; e.g., [www.corpora.aifdb.org/US2016](http://www.corpora.aifdb.org/US2016) for the full US2016 corpus. The online environment makes it possible to download the corpus, and to access the tools interacting with the Argument Web (Bex et al. 2013). The Argument Web is a vision of inter-connected argumentative content produced and manipulated through an online infrastructure of computational tools that facilitate interaction with argumentative content in various ways (Rahwan et al. 2007). Figure 4 shows a small part of the US2016 corpus the AIFdb Corpora online interface, which allows downloads in various file formats (indicated by label “1” in the figure), shows miniature diagrammatic overviews of argumentative structures (label “2”), and provides direct access to, for example, the Argument Analytics module (Lawrence et al. 2016, 2017) to explore the quantitative characteristics and metrics of the corpus (“3”), and the aforementioned OVA to manipulate the annotation (“4”).

### 4.3 Validation

To validate corpus annotation, pairwise inter-annotator agreement values are calculated for both the television sub-corpus (US2016tv) and the Reddit sub-corpus (US2016reddit). A sample of each corpus was annotated by four annotators (A1, A2, A3 and A4) in the case of the televised debates and two annotators (A3 and A4) for the Reddit discussions.

For the US2016tv corpus, comprised of US2016R1tv, US2016D1tv and US2016G1tv, a 10.5% (word count) sample was selected for inter-annotator agreement calculation purposes. This sample was selected on the basis that a) all annotators involved in the annotation process of a sub-corpus (e.g. US2016D1tv) must be compared to each other, and b) the total of all randomly selected excerpts must equal or exceed 10% of the sub-corpus. In the case of the US2016reddit corpus, a sample was selected for inter-annotator agreement encompassing every tenth argument map until at least 10% (word count) of the original corpus size was achieved comparing both annotators resulting in a total sub-sample of 12.6%.

In Table 2, agreement results are reported in terms of Cohen's  $\kappa$  scores (Cohen 1960). On the basis of a pairwise comparison between annotators, and normalising for word count, the combination of the television and Reddit debates gives an overall Cohen's  $\kappa$  of 0.610. Landis and Koch (1977) interpret  $\kappa$ -scores of 0.41–0.60 as moderate agreement,  $\kappa$ -scores of 0.61–0.80 as substantial agreement and  $\kappa$ -scores of 0.81–1.00 as almost perfect agreement. The achieved substantial agreement falls within upper expectation for the argumentation annotation task due to the great number and variety of relations and possible interpretations available to the annotators—especially when compared to other tasks such as named-entity recognition or part-of-speech tagging which are expected to achieve almost perfect agreement.



**Table 2** Cohen's  $\kappa$  and CASS- $\kappa$  scores for television debates and Reddit discussions normalised by overall corpus word count

Annotator pair	Data source	Sample word count	Cohen's $\kappa$	CASS- $\kappa$
A1–A2	US2016tv	1890	0.453	0.624
A1–A3	US2016tv	1318	0.658	0.776
A1–A4	US2016tv	275	0.540	0.670
A2–A3	US2016tv	1380	0.387	0.594
A2–A4	US2016tv	925	0.624	0.790
A3–A4	US2016tv	407	0.710	0.810
A3–A4	US2016reddit	4920	0.693	0.817
<b>Combined (US2016)</b>		<b>11,115</b>	<b>0.610</b>	<b>0.752</b>

**Table 3** Intermediate CASS scores for partial annotation tasks on television debates and Reddit discussions, in terms of Fournier and Inkpen's segmentation similarity statistic S and Cohen's  $\kappa$ 

Corpus	Segmentation	Transitions	Illocutionary connections	Propositional relations
<b>US2016tv</b>	0.907	0.559	0.491	0.702
<b>US2016reddit</b>	0.958	0.810	0.645	0.817

While we prefer Cohen's  $\kappa$  metric over percentage agreement, because it accounts for chance agreements between annotators, it has the drawback that errors can be passed from text segmentation—a non-fixed task—to identifying relations, thus not providing a comprehensive agreement score. Duthie et al. (2016b) introduce the Combined Argument Similarity Score  $\kappa$  (CASS- $\kappa$ ) aimed at overcoming this by calculating intermediate agreement scores for the composite tasks of text segmentation, annotation of dialogical relations, and annotation of propositional relations, before combining the detailed calculations into an overall CASS- $\kappa$  score (while still accounting for chance agreements.) In Table 2, we include the pairwise CASS- $\kappa$  scores, resulting in an overall CASS- $\kappa$  for US2016 of 0.752.

The intermediate CASS- $\kappa$  scores are only informative relative to the other intermediate CASS- $\kappa$  scores, as the annotation sub-tasks are not fully independent (see Sect. 3.2) and the scores cannot be normalised with respect to a shared unit of quantity, because the sub-tasks reference different units (locutions, propositions, etc.) not directly related to word counts. Nevertheless, they do give some insight into which parts of the annotation are relatively more difficult. The intermediate CASS scores in Table 3 show that the annotation of illocutionary connections turns out to be more challenging than that of propositional relations, discourse transitions, and segmentation (all recorded as Cohen's  $\kappa$  values, except for segmentation, which is calculated in terms of Fournier and Inkpen (2012)'s S metric for segmentation similarity). The difficulty of annotating illocutionary connections is not surprising, as Budzynska et al. (2016) previously observed that the closeness between certain types of illocutionary connections can make them difficult to distinguish.



Examples (4) and (5) are two cases in point.<sup>9</sup> In Example (4), one annotator analysed moderator Megyn Kelly's utterance as comprising an *Assertive Challenge* where the other went for *Assertive Question*. These types of illocutionary connection share many characteristics—syntactically questions, both carrying assertive force—but differ in the burden of proof they allot to the addressee: a question is a request for explanation, whereas a challenge prompts supporting argumentation. Example (4) was annotated as *Pure Question* once and as *Assertive Question* by the second annotator. Again, these illocutionary connections share the surface form of a question, but in this case they differ in the assertive force conveyed.

- (4) Megyn Kelly: *Does that sound to you like the temperament of a man we should elect as president[...]?*
- (5) doktorphil: *How are you going to accomplish all these lofty, ridiculous goals?*

Tables 2 and 3 show that the inter-annotator agreement is generally higher for US2016reddit than for US2016tv. While further qualitative error analysis is required to find a precise explanation for the difference, we hypothesise that, aside from the individual annotators involved, the main factor is to be found in the discourse cohesion that is present in the Reddit threads. The interface of the Reddit forum is such that the shorter dialogue turns, and explicit response-structure between posts within one thread, make it easier to identify discourse transitions, and relations between propositions that are temporally further apart ('long-distance' relations). Furthermore, the smaller contiguous blocks of text in the Reddit corpus make annotation less exhausting, leading to fewer annotation errors—something we have since aimed to address for longer contiguous texts by redesigning the annotation task as an iterative process implementing various stages of gate-keeping and error-checking between annotators (Budzynska et al. 2018).

#### 4.4 Corpus properties

The combined US2016 corpus comprises 97,999 words (tokens). The annotated television debates account for 58,900 words, and the online reactions on Reddit for 39,099 words. To the best of our knowledge, this makes US2016 the largest corpus of argumentative dialogue annotated to this detail that is currently available. Most publicly available corpora of equal or larger size that we are aware of are monological and are annotated on the basis of more lightweight theoretical models of argumentation.

In addition to word count, we propose the 'argument density' of a corpus as a comparative measure. We calculate argument density by normalising the number of annotated inference relations to the word count of the corpus. The argument density of US2016 is 0.028 (meaning that there is one annotated inference relation for every 36 words), the US2016tv sub-corpus scores 0.026, and US2016reddit 0.031. These

<sup>9</sup> Example (4)—taken from our annotation of the first Republican primaries television debate on 6 August 2015 in Cleveland, Ohio—is available online at [www.aifdb.org/argview/10450](http://www.aifdb.org/argview/10450) and [www.aifdb.org/argview/10470](http://www.aifdb.org/argview/10470). Example (5)—taken from our annotation of the Reddit reaction to the first Republican primaries television debate on 6 August 2015 in Cleveland, Ohio—is available online at [www.aifdb.org/argview/10394](http://www.aifdb.org/argview/10394) and [www.aifdb.org/argview/10535](http://www.aifdb.org/argview/10535).

scores are similar to those of other corpora in AIFdb: AraucariaDB (a compilation of Araucaria<sup>10</sup> analyses) with 0.028 for 80,000 words (Reed et al. 2008a); MM2012c (analyses of episodes of BBC Radio 4's Moral Maze program) with 0.022 for 39,694 words; DMC (a corpus of dispute mediation) with 0.033 for 28,956 words (Janier and Reed 2016).

The largest annotated corpus of monological argumentative discourse that we are aware of is the second version of the Argument Annotated Essays corpus (Stab and Gurevych 2017) (AAEC2), comprising 131,633 words (according to our count). If we calculate the argument density of the AAEC2 corpus by dividing the number of annotated support relations by the number of words, we get a very similar score of 0.027. While this could provide a useful comparative measure, it is hard to determine what exactly it signifies, because the annotation schemes of the various corpora are based on different underlying conceptualisations of argumentative concepts. This means that there is no guarantee that we are comparing actually comparable properties of the corpora. As more corpora start using a common representation standard (or conceptual ontology), such as the AIF<sup>+</sup> ontology (Chesñevar et al. 2006; Reed et al. 2008b) underlying AIFdb, this obstacle can be overcome. The AIF<sup>+</sup> ontology can function as an interlingua for the expression of different theoretical conceptualisations of argumentatively relevant notions, thus enabling the comparison of various corpora based on a common notion of argument density.

Table 4 presents the number of propositions and argumentatively relevant relations between them. This includes the number of inference, conflict, and rephrase relations between the propositional contents of the segmented locutions for each of the sub-corpora. The full US2016 corpus contains 4197 annotated argumentatively relevant relations between propositions: 2754 instances of inference, 823 conflicts between propositions, and 620 rephrases. Identical propositions that occur more than once in a corpus (for example because the duplicates occur in two of the constitutive sub-corpora of a collated corpus) are only counted once, i.e. the counts are of types not tokens.

The texts annotated as part of US2016 are segmented into a total of 8937 locutions: 4671 in the television debate corpus US2016tv, and 4266 in the corresponding Reddit discussion corpus US2016reddit. In Table 5, we give an overview of the dialogical properties of US2016. The corpus contains 12,965 occurrences of the various types of illocutionary connection between locutions and propositions and between or transitions and propositional relations. The most common illocutionary connection is *Asserting* with 7886 instances, followed by the quintessential illocutionary connection for an argumentation corpus, *Arguing*, which occurs 2714 times in US2016. The three sub-types of *Questioning* (*Pure*, *Assertive* and *Rhetorical*) together add up to 590 instances. *Disagreeing* (776 instances) turns out to be much more common than *Agreeing* (214 instances). The category 'Other' contains all less common types of illocutionary connection, such as 19 counts of *Contradicting* and a combined total of 67 occurrences of the three sub-types of *Challenging* (*Pure*, *Assertive*, and *Rhetorical*).

<sup>10</sup> Araucaria (Reed and Rowe 2004) is a popular, early argument diagramming software tool, which can be seen as a precursor to the OVA software used in the annotation of the US2016 corpus.

**Table 4** Proposition and propositional relation counts for the US2016 corpora

Corpus	Propositions	Inference	Conflict	Rephrase
US2016R1tv	1368	482	61	88
US2016R1reddit	1173	389	154	88
<b>US2016R1</b>	<b>2540</b>	<b>871</b>	<b>215</b>	<b>176</b>
US2016D1tv	1439	564	54	105
US2016D1reddit	1378	437	233	89
<b>US2016D1</b>	<b>2816</b>	<b>1001</b>	<b>287</b>	<b>194</b>
US2016G1tv	1473	505	79	140
US2016G1reddit	1279	377	242	110
<b>US2016G1</b>	<b>2752</b>	<b>882</b>	<b>321</b>	<b>250</b>
US2016tv	4277	1551	194	333
US2016reddit	3827	1203	629	287
<b>US2016</b>	<b>8099</b>	<b>2754</b>	<b>823</b>	<b>620</b>

The Argument Analytics module can be used to observe various characteristics of the two genres covered by the US2016 corpus, pointing at the similarities and differences in the discourse dynamics characteristic of debates on live television and on online discussion forums. For example, in Fig. 5 we can see that the most frequent annotated relation between propositions is that of inference, indicating the support of one proposition for the acceptability of another. This predominance of inference over conflict and rephrase is a constant throughout the US2016 corpus, although the relative proportions vary (ranging from 70 to 78% for the television debates, and from 52 to 62% for the Reddit discussions).

Aside from the difference in the proportion of inference relations, there is a clear distinction between the television debates and the Reddit discussion when we consider the conflict and rephrase relations. As Fig. 5 shows, the proportion of rephrases is higher in the television debates at 14–19% than the proportion of conflicts at 7–11%. In contrast, in the three Reddit discussion sub-corpora, we observe the reverse pattern, with only 12–15% of the annotated propositional relations consisting of rephrases and 24–33% consisting of conflicts. Overall, Fig. 5 shows that the relative proportion of conflict is greater in the Reddit discussion, whereas rephrases are used proportionally more often in the television debates. Inferential relations constitute more than half of the propositional relations across the board, but they occur more often in the television debates, while the Reddit discussions contain proportionally more explicit conflict.

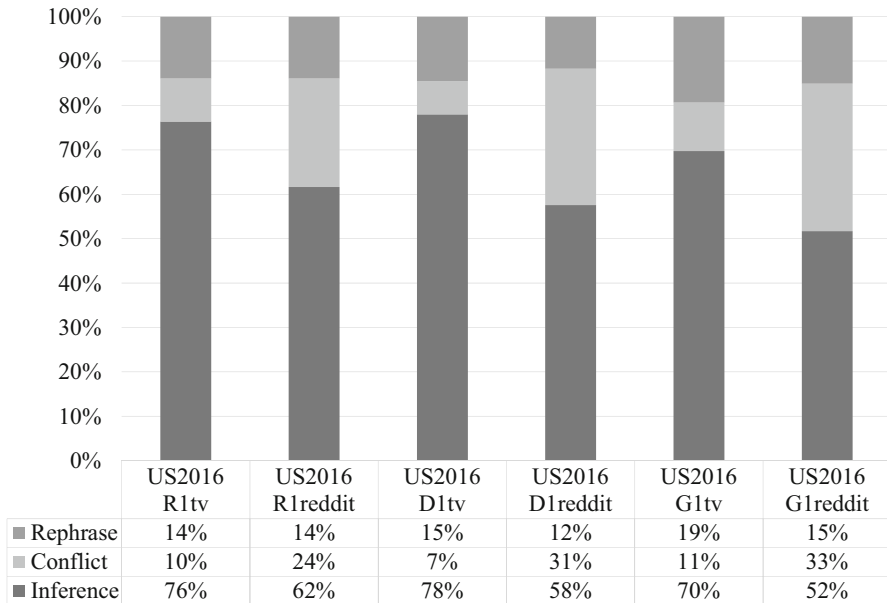
## 5 Intertextual correspondence between television debates and social media discussions

### 5.1 Intertextual correspondence

Thus far, we have treated the sub-corpora of the television debates and the Reddit reactions as two independent corpora, together constituting the US2016 corpus

**Table 5** Locution and illocutionary connection counts for the US2016 corpora

Corpus	Locations	Agreeing	Arguing	Asserting	Default illocuting	Disagreeing	Questioning	Restating	Other
US2016R1tv	1521	16	471	1363	29	45	96	60	35
US2016R1reddit	1323	62	387	1134	27	155	100	63	22
<b>US2016R1</b>	<b>2844</b>	<b>78</b>	<b>858</b>	<b>2497</b>	<b>56</b>	<b>200</b>	<b>196</b>	<b>123</b>	<b>57</b>
US2016D1tv	1566	13	545	1386	47	46	109	62	9
US2016D1reddit	1523	79	428	1321	35	231	83	62	22
<b>US2016D1</b>	<b>3089</b>	<b>92</b>	<b>973</b>	<b>2707</b>	<b>82</b>	<b>277</b>	<b>192</b>	<b>124</b>	<b>31</b>
US2016G1tv	1584	7	507	1465	17	62	76	121	30
US2016G1reddit	1420	37	376	1217	46	237	126	65	33
<b>US2016G1</b>	<b>3004</b>	<b>44</b>	<b>883</b>	<b>2682</b>	<b>63</b>	<b>299</b>	<b>202</b>	<b>186</b>	<b>63</b>
US2016iv	4671	36	1523	4214	93	153	281	243	74
US2016reddit	4266	178	1191	3672	108	623	309	190	77
<b>US2016</b>	<b>8937</b>	<b>214</b>	<b>2714</b>	<b>7886</b>	<b>201</b>	<b>776</b>	<b>590</b>	<b>433</b>	<b>151</b>



**Fig. 5** Distribution of propositional relations in sub-corpora of US2016

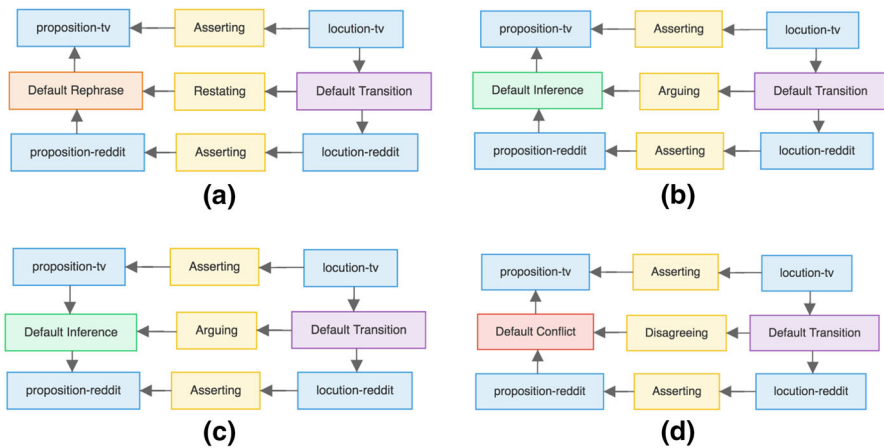
based on topical and temporal relatedness. In some sense, this is unproblematic: the two sub-corpora can be considered as independent, each with their own value. Our claim, however, is that their value can be transformed by exploring the connection between the television debates and the online reaction. To this avail, we extend the annotation to capture ‘intertextual correspondence’: the topical interrelatedness between the contents of independent text corpora (Visser et al. 2018a).<sup>11</sup>

Because the online discussion on Reddit is a direct reaction to the candidates’ election debates on live television and both are examples of highly persuasive communicative contexts, there is a richness of argumentative connectivity to explore. Some contributors on Reddit will, for example, draw conclusions on the basis of the arguments presented in the television debates. Others will voice their disagreement or rephrase the candidates’ utterances. The annotation of intertextual correspondence as part of the US2016 annotation project epitomises the vision of the Argument Web by enabling the interconnection of argumentative content in separate corpora and from different communicative domains and genres.

## 5.2 Annotation of intertextual correspondence

The annotation of the intertextual correspondence between the television debates and the Reddit discussions follows the general principles set out as part of the

<sup>11</sup> We use the term ‘intertextual’ due to the resemblance of the correspondence between corpora to the postmodern idea that texts can only be properly understood in their relation to the larger body of extant texts (Kristeva 1977).



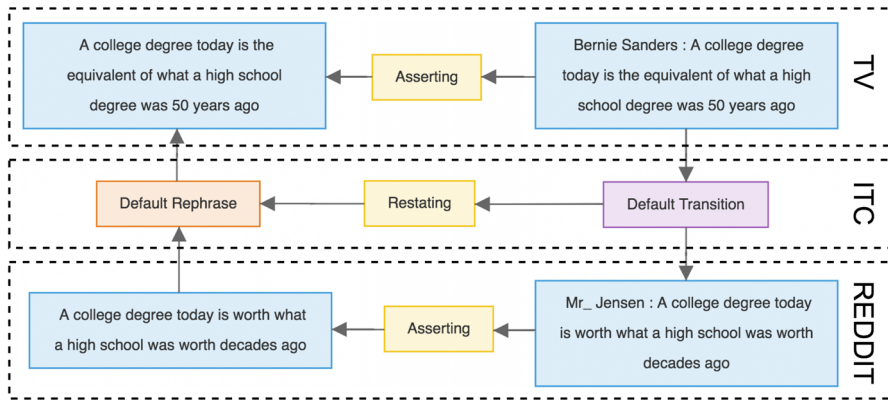
**Fig. 6** Diagrammatic visualisation of the four common patterns of intertextual correspondence

annotation guidelines in Sect. 3.2. No new locutions or propositions are created as part of the annotation of intertextual correspondence, because no new text excerpts are introduced to the corpora. By establishing connections between the existing television and Reddit sub-corpora, the annotation of intertextual correspondence only creates new transitions, illocutionary connections and propositional relations.

One starting point for the annotation is that Reddit contributors can respond to the candidates' utterances in the television debates, but never vice versa. This means that the flow of the 'intertextual dialogue' always goes from television debate to Reddit commentary. In other words, the implicit dialogue protocol that is followed is such that transitions between locutions only go from a locution part of US2016tv to a locution part of US2016reddit.

Based on the contextual characteristics of the two genres of television debates and social media discussion, four common annotative patterns can be expected to occur most frequently (although variations are possible). To make it easier to discuss the four patterns, we will use the suffixes '-tv', '-reddit', and '-itc' when we refer to the elements of the annotations that are part of, respectively, the television debate sub-corpus (US2016tv), the Reddit discussion sub-corpus (US2016reddit), and the intertextual correspondence sub-corpus (US2016itc).

The first common pattern, visualised in Fig. 6a, deals with rephrases on Reddit of what is said in the television debates. Whether introduced directly or by means of reported speech, the politician's (or moderator's) statement is often not literally repeated, but rather reformulated to some degree, introducing an intertextual rephrase relation. This results in an annotation structure where the middle row of three nodes connect content from the US2016reddit sub-corpus directly to the US2016tv sub-corpus, by means of a rephrase-itc node which is anchored through a restating-itc node in a transition-itc node. An example instantiating this pattern of intertextual annotation of the direct restating on Reddit of content from the television debates is visualised in Fig. 7. On the top we see the trinity of



**Fig. 7** Diagrammatic visualisation of an intertextual rephrase relation, with an indication of the parts coming from US2016tv, US2016reddit, and US2016itc

proposition-tv, illocution-tv and locution-tv as part of the US2016D1tv sub-corpus: in this case, a claim about college affordability by Bernie Sanders (then candidate for the Democratic nomination). On the bottom, Fig. 7 contains the associated proposition-reddit, illocution-reddit and locution-reddit that are part of the US2016D1reddit sub-corpus: Mr\_Jensen's reformulation as part of a discussion about what exactly Sanders meant. The middle row, the 'intertextual layer', contains three new relations, inserted as part of the annotation of intertextual correspondence. The transition-itc leading from locution-tv to locution-reddit shows that the comment on Reddit by Mr\_Jensen is a dialogical continuation of what was said by Sanders in the television debate. The rephrase-itc relation from proposition-reddit to proposition-tv reflects the intertextual rephrase. Lastly, the illocution-itc of *Restating* anchors the *Default Rephrase* relation between the two propositions in the *Default Transition* between the two locutions.

The second pattern occurs when a Reddit user argues why what was said in the television debate should be accepted. In this case, the Reddit user provides an argument in defence of the acceptability of a proposition (or locution) advanced on television. The resulting annotation pattern is visualised in Fig. 6b. The pattern is similar to that of Fig. 6a, with the rephrase-itc replaced by an inference-itc and the illocution-itc changed to *Arguing*.

The third pattern is closely related to the second, but reverses the direction of the inference relation. Instead of arguing why something in the television debate is acceptable, the Reddit user draws a conclusion on the basis of what was said in the television debate. The illocutionary connection *Arguing* anchors a relation of inference-itc going from a proposition-tv to a proposition-reddit. The resulting annotation pattern is visualised in Fig. 6c, and differs from Fig. 6b only in the direction of the inference-itc, which is reversed.

The fourth pattern concerns disagreement, rather than rephrase or the drawing of conclusions or providing of additional reasons. Voicing opposition to what was asserted on television results in the structure of Fig. 6d: an illocution-itc

**Table 6** Overview of composition of the US2016\* corpus including intertextual correspondence

US2016*	US2016R1*	US2016D1*	US2016G1*
US2016tv	US2016R1tv	US2016D1tv	US2016G1tv
US2016reddit	US2016R1reddit	US2016D1reddit	US2016G1reddit
US2016itc	US2016R1itc	US2016D1itc	US2016G1itc

*Disagreeing* is introduced, anchoring a conflict-itc between proposition-reddit and proposition-tv (if the disagreement is with the acceptability of the content of what was stated on television)—in exceptional cases, the conflict-itc might target the locution-tv: if the opposition is directed not at the content, but rather at the locutive act itself, i.e. the acceptability of the speech act performed.

### 5.3 The intertextual correspondence sub-corpus

To position the intertextual correspondence sub-corpus within the full US2016 corpus, we can revise Table 1 to include the US2016itc sub-corpus. Table 6 shows how US2016itc fits into the existing composition of the corpus (introducing the suffix ‘itc’ to name corpora dedicated to intertextual correspondence). Following the same pattern as before, three new sub-corpora, US2016R1itc, US2016D1itc, and US2016G1itc, contain the annotations of intertextual correspondence for respectively the first Republican primaries debate, the first Democratic primaries debate, and the first general election debate. A derived sub-corpus US2016itc collates all intertextual correspondence annotations. The four new sub-corpora can be accessed on AIFdb Corpora as outlined before—e.g., [www.corpora.aifdb.org/US2016itc](http://www.corpora.aifdb.org/US2016itc) for the sub-corpus containing all intertextual correspondence annotation.

The addition of intertextual correspondence annotations does not affect the US2016tv and US2016reddit sub-corpora, but it is incorporated into the collated cross-genre sub-corpora for the individual debates. We indicate the inclusion of intertextual correspondence in these sub-corpora by adding ‘\*’ to the corpus identifiers. The same holds for the main US2016 corpus, which, after the extension with US2016itc, is identified as US2016\* in Table 6.

Without adding new locutions or propositions to the corpus, the annotation of intertextual correspondence enriches the US2016 corpus by making explicit the relations between the television and Reddit sub-corpora. For this reason, Table 7 does not include counts of words, locutions or propositions, but rather the increases in the propositional relations of inference, conflict and rephrase, and the corresponding illocutionary connections of *Arguing*, *Disagreeing* and *Restating* by means of which they are stereotypically anchored. In total, the annotation of intertextual correspondence adds 339 propositional relations to the US2016 corpus (an increase of 8%) and 366 illocutionary connections (an increase of 3%). The strongest effect is found on the rephrase relations (144 in US2016itc), which is not surprising as this indicates that contributors on Reddit restate what the candidates



**Table 7** Propositional relation and illocutionary connection counts for the US2016itc corpus

Corpus	Inference	Conflict	Rephrase	Arguing	Disagreeing	Restating	Other
US2016R1itc	19	48	55	19	52	55	4
US2016D1itc	38	45	69	36	53	68	14
US2016G1itc	19	26	20	19	26	20	0
US2016itc	76	119	144	74	131	143	18
US2016	2754	823	620	2714	776	433	9042
US2016*	2830	942	764	2788	907	576	9060

said during the television debates as input for their own online discussion. The addition of 76 inference relations marginally raises the argument density of the US2016\* corpus to 0.029. The intertextual correspondence for the first general election debate (compiled in US2016G1itc) exhibits relatively low numbers of conflict and inference, while that for the first Democratic primary debate (US2016D1itc) contains relatively many inferences.

## 6 The US2016 corpus as a resource

Not precluding other uses, the US2016 corpus was developed with two main applications in mind: as a resource for argument mining and for quantitative empirical studies. With respect to the first application, US2016 is developed in such a way that it can serve as a resource in the development of reliable automated annotation methods for argumentative discourse. The automated retrieval of argumentative structures from natural language text is commonly referred to as argument(ation) mining (Palau and Moens 2009). Just like the related research on sentiment analysis and opinion mining (Pang and Lee 2008), many argument mining techniques are based on machine learning.

A requirement for the development of successful machine learning algorithms is the availability of annotated data. The quality of the algorithm's output depends on both the quantity and quality of the data used as input. The US2016 corpus is one of the largest corpora of its kind, and is annotated to a high level of detail, covering dialogical structures, illocutionary connections, and relations of inference, conflict and rephrase between propositions, in addition to the segmentation into argumentative discourse units. Furthermore, US2016 combines heterogeneous data from two genres—televised election debates and social media discussions—including the annotated intertextual correspondence between them.

The properties that make US2016 suitable as a resource for machine learning approaches to argument mining should also prove valuable for natural language processing in general, and in the study of non-argumentative linguistic phenomena, such as question-answering, the protocol of dialogue, or the genre and context specific preconditions of communication. For example, in an ongoing project,

Koszowy extends the annotation of the US2016 corpus with rhetorical aspects in order to study the role of ethos in persuasive discourse, continuing on previous work on other corpora (Koszowy and Budzynska 2016) and working towards the mining of ethotic structures (Duthie et al. 2016a). As Lawrence and Reed (2017) show, the US2016 corpus can also be employed in argument mining techniques not based on machine learning. The authors exploit the interconnected graph data structure of the US2016G1tv sub-corpus by calculating centrality and divisiveness scores of propositions to reconstruct the structure of the argumentation.

The second intended application of the US2016 corpus is to provide quantitative means to empirically study the properties of actual argumentative discourse. The corpus can facilitate the empirical testing of argumentation theoretical hypotheses on the basis of large scale quantitative data about argumentation in practice—to answer questions like: how frequently is argumentation in practice signalled with an explicit lexical indicator, such as “because”?

The annotations, the intertextual relation between debate and online discussion, and the graph-based methods of extracting global metrics (with the Argument Analytics module of the Argument Web mentioned in Sect. 4.2) provide data that may be of use in political science studies and Critical Discourse Analysis (Fairclough 1995). Within an educational setting, the searchable US2016 annotations can be used to retrieve relatable examples from actual argumentative practice for use in critical thinking and debating classes.

Moving from the academic realm to societal applications, Argument Analytics provides an online interface to the quantitative characteristics of corpora, like those we described in Sect. 4.4. Insight into the structure and properties of the argumentation can contribute to data-driven rather than interpretative sense-making and decision-making in the public domain (Lawrence et al. 2017). A better understanding of the candidates’ positions and reasoning in a televised election debate can further public engagement with the issues and contribute to a well-informed electoral vote. Such deeper insight into the dynamics and structure of the argumentative interaction can also be valuable for the involved politicians and their campaign teams.

An important next step would be to provide such Argument Analytics live, e.g., during a television debate. Near-realtime annotation (piloted for an episode of the BBC 4 Radio programme Moral Maze [www.arg.tech/wall](http://www.arg.tech/wall)) would allow Argument Analytics to be accessed during the debate making it easier for the public at large to understand the argumentative proceedings in detail when it’s most relevant, and supplying pundits and political analysts with quantitative empirical data to support their running commentary.

## 7 Related work

To the best of our knowledge, the US2016 corpus we introduce in the current paper, is the first publicly available corpus with integrated annotation of argumentatively relevant propositional structures and pragmatic annotations of dialogical relations. Furthermore, the corpus is unique in its application of such detailed annotations on combined data from two distinct but related types of communicative activity:

television debates and social media discussions. As a result, existing related work comes from various areas, such as annotated corpora of argumentation, pragmatically annotated corpora, resources of political discourse, and social media studies.

*Argumentation.* Existing resources of argumentative discourse include those created for the purpose of automated argument reconstruction, such as the annotation scheme for Argumentative Zoning (Teufel et al. 1999) and its elaboration for academic texts (Teufel et al. 2009). The continued progress in argument mining (Palau and Moens 2009) increases the need for annotated corpora, leading to, e.g., the development of a corpus of argumentative microtexts in both German and English (Peldszus and Stede 2015), work on online user comments (Park and Cardie 2014), the Internet Argument Corpus 2.0 (Abbott et al. 2016), as well as several corpora stored in AIFdb annotated only with propositional argument structures, such as Regulation Room Divisiveness (Konat et al. 2016) and AraucariaDB (Reed et al. 2008a). Noteworthy is also the collection of Darmstadt Corpora, which covers various genres, such as persuasive essays, scientific papers, news articles, and online discourse (Stab and Gurevych 2017). These resources are focussed on the propositional dimension of argumentation, and tend to disregard the conversational dialogue genres.

Like the US2016 corpus, some existing corpora combine the annotation of dialogue structures, illocutionary connections and propositional relations, thus combining pragmatic (or dialogical) and inferential (or propositional) annotations (Reed 2006). The US2016 corpus is the latest in a series of corpora of argumentative texts annotated on the basis of Inference Anchoring Theory (see Sect. 3.1). Other such corpora include the MM2012 corpus of BBC Radio 4 Moral Maze programs (Budzynska et al. 2014), and the Dispute Mediation Corpus (Janier and Reed 2016). While not resulting in one integrative annotation, Stede et al. (2016) report on the multi-layer annotation of the texts of the argumentative microtext corpus (Peldszus and Stede 2015) on the basis of Rhetorical Structure Theory (Mann and Thompson 1988) and Segmented Discourse Representation Theory (Asher and Lascarides 2003), thus also combining pragmatic with argumentative annotation of the same texts.

*Pragmatics.* Moving away from the focus on argumentation, modern pragmatics relies on large annotated corpora of conversational data (Romero-Trillo 2017). Pragmatic annotation can cover various facets of discourse, ranging from, e.g., dialogue acts (Weisser 2014, 2016; Vail and Boyer 2014) to the discourse semantics of the Penn Discourse Treebank (Prasad et al. 2014). Rhetorical Structure Theory (RST) (Mann and Thompson 1988) provides a foundation for several annotated corpora, including the RST Treebank (Carlson et al. 2002), as well as purpose-built corpora, such as the analysis by Das and Taboada (2017) of indicators of coherence relations. Berzlánovich and Redeker (2012) analyse the interaction between genre and coherence relations in another study based on RST. The RST-based corpora can also be employed for automated RST parsing, as shown by Feng and Hirst (2012). A broad range of pragmatic features, such as speech acts, tone movements, discourse markers, utterance tags and quotatives, are annotated by Kirk (2016) on the basis of the Pragmatic Annotation Scheme developed for the SPICE-Ireland Corpus.

*Political discourse.* Corpus-based studies are also found in the political field. Laver et al. (2003), for example, employ a quantitative method for extracting policy

positions from political texts. The communicative aspects of the 2016 US presidential elections are also approached from a political sciences perspective, as evidenced by, e.g., the in-depth study by Wells et al. (2016) of Trump's hybrid media campaigning, analysing how the television debates and other events influence the coverage the candidates receive in the traditional media and on social media. Emphasising the role that the medium plays in shaping the political discourse, Giltrow and Stein (2009) show how identifying strategies and arguments can be used to determine the goals and values of politicians.

*Social media.* Sharing our focus on the interaction between user-generated social media content and argumentation, Walker et al. (2012b) compile large corpora, which are subsequently used in various annotation projects on, e.g., disagreement (Abbott et al. 2011), and stance-taking (Walker et al. 2012a). Mullen and Malouf (2006) also focus on online informal political discourse and explore the application of sentiment analysis techniques to this communicative context. Social media activity can serve as a predictor of voting preferences. For example, more Tweets can result in more votes (DiGrazia et al. 2013), and Twitter data enable the classification of users as Democrats or as Republicans based on the political content shared (Colleoni et al. 2014).

Motivated by these results, several studies focus on the 2016 US presidential election campaigns. These studies provide a wider context to our US2016 corpus, without being similar resources in terms of what is annotated. One of the election-related datasets (incidentally also called 'US2016') contains details about the Twitter followers of Clinton and Trump; including the number of followers of each candidate, their geographical location, the number of their own followers, and their profile images. Such data can, in turn, be used to derive further information: the followers' images, for example, can be used to determine their gender and race (Wang et al. 2016a). In a follow-up study on this dataset, Wang et al. (2016b) perform a topic analysis of Trump's followers on Twitter, looking for the correlation between the topic and the number of 'likes' each message attracts—finding that the most favoured topic for the Trump followers is attacking the Democrats.

While Reddit is less commonly used as a data source than, e.g., Twitter, Facebook or Amazon, our US2016 corpus is by no means the first to include Reddit material. Gao (2016) uses Reddit data to visualise opinion clustering, and studies by Wei et al. (2016) and Tan et al. (2016) look at more argumentative and persuasive aspects of Reddit discussions. Whatever social media platform the data is sourced from, the predictive value is, of course, not guaranteed, as evidenced by the study by Bovet et al. (2016) which came to election outcome prognoses consistent with traditional opinion polls—which, as we now know, were also largely wrong.

## 8 Conclusion

The US2016 corpus and its component parts are a unique set of resources that represents a number of firsts. US2016tv is the largest corpus of analysed dialogical argumentation currently available. As a whole, this is the largest corpus annotated

according to argumentation theoretic principles. The inclusion of US2016itc delivers for the first time cross-corpus connections that not only make US2016 unique but demonstrates the way in which intertextual correspondence analysis might be used in general to extend the value of extant corpora.

Of course, the value of any resource ultimately lies in the uses to which it can be put: initial work with US2016 has demonstrated its utility in two domains. First, by providing the raw data for subsequent processing, the corpus allows evidence-based analysis of debates at scale for the first time. This work is being further pursued to deliver a broad range of analytics that can deliver insight and summary of extended argumentative debates. The second domain in which US2016 is being exploited is argument mining: acting as training data for machine learning techniques and as gold standard targets against which to test. As resources made available freely to the academic community in perpetuity, the goal is that the US2016 corpora should add significantly to the research programmes in both of these exciting, high-growth areas.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Abbott, R., Ecker, B., Anand, P., & Walker, M. (2016). Internet Argument Corpus 2.0: An SQL schema for dialogic social media and the corpora to go with it. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.) *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA).
- Abbott, R., Walker, M., Anand, P., Fox Tree, J. E., Bowmani, R., & King, J. (2011). How can you say such things?!?: Recognizing disagreement in informal political argument. In *Proceedings of the workshop on Language in Social Media (LSM)*, (pp. 2–11). Association for Computational Linguistics.
- Asher, N., & Lascarides, A. (2003). *Logics of conversation*. Cambridge: Cambridge University Press.
- Austin, J. L. (1962). *How to do things with words*. Oxford: Clarendon Press.
- Berzlánovich, I., & Redeker, G. (2012). Genre-dependent interaction of coherence and lexical cohesion in written discourse. *Corpus Linguistics and Linguistic Theory*, 8(1), 183–208.
- Bex, F., Lawrence, J., Snaith, M., & Reed, C. (2013). Implementing the argument web. *Communications of the ACM*, 56(10), 66–73.
- Bovet, A., Morone, F., & Makse, H. A. (2016). Predicting election trends with Twitter: Hillary Clinton versus Donald Trump. CoRR abs/1610.01587.
- Budzynska, K., Janier, M., Reed, C., & Saint-Dizier, P. (2016). Theoretical foundations for illocutionary structure parsing. *Argument and Computation*, 7(1), 91–108. <https://doi.org/10.3233/AAC-160005>.
- Budzynska, K., Janier, M., Reed, C., Saint-Dizier, P., Stede, M., & Yaskorska, O. (2014). A model for processing illocutionary structures and argumentation in debates. In *LREC*, (pp. 917–924).
- Budzynska, K., Pereira-Fariña, M., De Franco, D., Duthie, R., Franco-Guillen, N., Hautli-Janisz, A., Janier, M., Koszowy, M., Marinho, L., Musi, E., Pease, A., Plüss, B., Reed, C., & Visser, J. (2018). Time-constrained multi-layer corpus creation. In *The 16th ArgDiaP Conference, Argumentation and Corpus Linguistics, ArgDiaP, Warsaw, Poland*, (pp. 1–6).
- Budzynska, K., & Reed, C. (2011). Whence inference. Tech. rep., University of Dundee.
- Carlson, L. (1983). *Dialogue games: An approach to discourse analysis*. Dordrecht: Kluwer.
- Carlson, L., Okunowski, M. E., & Marcu, D. (2002). *RST discourse treebank*. Philadelphia: Linguistic Data Consortium.

- Chesñevar, C., McGinnis, J., Modgil, S., Rahwan, I., Reed, C., Simari, G., et al. (2006). Towards an argument interchange format. *The Knowledge Engineering Review*, 21(04), 293–316.
- Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20(1), 37–46.
- Colleoni, E., Rozza, A., & Arvidsson, A. (2014). Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *Journal of Communication*, 64(2), 317–332.
- Das, D., & Taboada, M. (2017). RST signalling corpus: A corpus of signals of coherence relations. *Language Resources and Evaluation*, 52, 149.
- DiGrazia, J., McKelvey, K., Bollen, J., & Rojas, F. (2013). More tweets, more votes: Social media as a quantitative indicator of political behavior. *PLoS ONE*, 8(11), e79449.
- Duthie, R., Budzysnka, K., & Reed, C. (2016a). Mining ethos in political debate. In P. Baroni, M. Stede, T. Gordon (Eds.) *Proceedings of the sixth international conference on computational models of argument (COMMA 2016)*, (pp. 299–310). IOS Press.
- Duthie, R., Lawrence, J., Budzysnka, K., & Reed, C. (2016b). The CASS technique for evaluating the performance of argument mining. In *Proceedings of the 3rd workshop on argument mining*, (pp. 40–49). Association for Computational Linguistics.
- Fairclough, N. (1995). *Critical discourse analysis: The critical study of language*. London: Longman.
- Feng, V. W., & Hirst, G. (2012). Text-level discourse parsing with rich linguistic features. In *Proceedings of the 50th annual meeting of the association for computational linguistics: Long papers*, (Vol. 1, pp. 60–68). Association for Computational Linguistics.
- Fournier, C., & Inkpen, D. (2012). Segmentation similarity and agreement. In *Proceedings of the 2012 conference of the North American chapter of the association for computational linguistics: Human language technologies*, (pp. 152–161). Association for Computational Linguistics.
- Gao, M. (2016). Intelligent interface for organizing online social opinions on Reddit. In *Companion publication of the 21st international conference on intelligent user interfaces, ACM, New York, NY, USA, IUI '16 Companion*, (pp 134–137). <https://doi.org/10.1145/2876456.2876464>.
- Gil de Zuniga, H., Garcia-Perdomo, V., & McGregor, S. C. (2015). What is second screening? Exploring motivations of second screen use and its effect on online political participation. *Journal of Communication*, 65(5), 793–815.
- Giltrow, J., & Stein, D. (2009). *Genres in the internet*. Issues in the Theory of Genre Amsterdam and Philadelphia: John Benjamins Publishing Company
- Givón, T. (1983). *Topic continuity in discourse: A quantitative cross-language study* (Vol. 3). Amsterdam: John Benjamins Publishing.
- Grimes, J. E. (1975). *The thread of discourse* (Vol. 207). Berlin: Walter de Gruyter.
- Grosz, B. J., & Sidner, C. L. (1986). Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3), 175–204.
- Hirschberg, J., & Litman, D. (1993). Empirical studies on the disambiguation of cue phrases. *Computational Linguistics*, 19(3), 501–530.
- Jacobs, S., & Jackson, S. (1982). Conversational argument: A discourse analytic approach. In J. R. Cox & C. A. Willard (Eds.), *Advances in argumentation theory and research* (pp. 205–237). Carbondale: Southern Illinois University Press.
- Janier, M., Lawrence, J., & Reed, C. (2014). OVA+: An argument analysis interface. In S. Parsons, N. Oren, C. Reed, F. Cerutti (Eds.) *Proceedings of the fifth international conference on computational models of argument (COMMA 2014)*, (pp. 463–464). Pitlochry: IOS Press.
- Janier, M., & Reed, C. (2016). Corpus resources for dispute mediation discourse. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis (Eds.) *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. Paris, France: European Language Resources Association (ELRA)
- Kirk, J. M. (2016). The pragmatic annotation scheme of the spice-ireland corpus. *International Journal of Corpus Linguistics*, 21(3), 299–322.
- Konat, B., Lawrence, J., Park, J., Budzysnka, K., & Reed, C. (2016). A corpus of argument networks: Using graph properties to analyse divisive issues. In *Language resources and evaluation conference (LREC 2016)*.
- Koszowy, M., & Budzysnka, K. (2016). Towards a model for ethotic structures in dialogical context. In P. Saint-Dizier, M. Stede (Eds.) *Proceedings of the COMMA 2016 workshop on Foundations of the Language of Argumentation*, (pp. 40–47).



- Kraus, S. (2013). *Televised presidential debates and public policy.*, Communication and society Milton Park: Taylor & Francis.
- Kristeva, J. (1977). Word, dialogue and novel. In L. S. Roudiez (Ed.), *Desire in language: A semiotic approach to literature and art* (pp. 64–91). Columbia: Columbia University Press.
- Landis, J., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 3, 159–174.
- Laver, M., Benoit, K., & Garry, J. (2003). Extracting policy positions from political texts using words as data. *American Political Science Review*, 97(2), 311–331. <https://doi.org/10.1017/S0003055403000698>.
- Lawrence, J., Bex, F., Reed, C., & Snaith, M. (2012). AIFdb: Infrastructure for the argument web. In *Proceedings of the fourth international conference on computational models of argument (COMMA 2012)*, (pp. 515–516).
- Lawrence, J., Duthie, R., Budzysnka, K., & Reed, C. (2016). Argument analytics. In P. Baroni, M. Stede & T. Gordon (Eds.) *Proceedings of the sixth international conference on computational models of argument (COMMA 2016)*, IOS Press.
- Lawrence, J., & Reed, C. (2014). AIFdb Corpora. In S. Parsons, N. Oren & C. Reed, F. Cerutti (Eds.) *Computational models of argument. Frontiers in artificial intelligence and applications*.
- Lawrence, J., & Reed, C. (2017). Using complex argumentative interactions to reconstruct the argumentative structure of large-scale debates. In I. Gurevych & I. Habernal (Eds.) *4th workshop on argument mining*.
- Lawrence, J., Snaith, M., Konat, B., Budzysnka, K., & Reed, C. (2017). Debating technology for dialogical argument: Sensemaking, engagement, and analytics. *ACM Transactions on Internet Technology*, 17(3), 24:1–24:23. <https://doi.org/10.1145/3007210>.
- Mann, W. C. (1988). Dialogue games: Conventions of human interaction. *Argumentation*, 2(4), 511–532.
- Mann, W. C., & Thompson, S. A. (1988). Rhetorical structure theory: Toward a functional theory of text organization. *Text-Interdisciplinary Journal for the Study of Discourse*, 8(3), 243–281.
- Mullen, T., & Malouf, R. (2006). A preliminary investigation into sentiment analysis of informal political discourse. In *AAAI symposium on computational approaches to analysing weblogs (AAAI-CAAW)*, (pp. 159–162).
- Palau, R.M., & Moens, M.F. (2009). Argumentation mining: The detection, classification and structure of arguments in text. In *Proceedings of the 12th international conference on artificial intelligence and law*, (pp. 98–107). ACM.
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1–2), 1–135. <https://doi.org/10.1561/15000000011>
- Park, J., & Cardie, C. (2014). Identifying appropriate support for propositions in online user comments. In *Proceedings of the first workshop on argumentation mining*, (pp. 29–38). Baltimore, MD: Association for Computational Linguistics. <http://www.aclweb.org/anthology/W/W14/W14-2105>
- Peldszus, A., & Stede, M. (2013). From argument diagrams to argumentation mining in texts: A survey. *International Journal of Cognitive Informatics and Natural Intelligence (IJCINI)*, 7(1), 1–31.
- Peldszus, A., & Stede, M. (2015). An annotated corpus of argumentative microtexts. In D. Mohammed & M. Lewiński (Eds.) *Argumentation and reasoned action. Proceedings of the 1st European conference on argumentation*, (pp. 801–816).
- Peters, G., & Woolley, J.T. (1999). The American presidency project. <http://www.presidency.ucsb.edu>. Accessed 11 Aug 2017.
- Peters, G., & Woolley, J.T. (2015a). Democratic candidates debate in Las Vegas, Nevada, October 13, 2015. <http://www.presidency.ucsb.edu/ws/?pid=110903>. Accessed 11 Aug 2017.
- Peters, G., & Woolley, J.T. (2015b). Republican candidates debate in Cleveland, Ohio, August 6, 2015. <http://www.presidency.ucsb.edu/ws/?pid=110489>. Accessed 11 Aug 2017.
- Peters, G., & Woolley, J.T. (2016). Presidential debate at Hofstra University in Hempstead, New York, September 26, 2016. <http://www.presidency.ucsb.edu/ws/?pid=118971>. Accessed 11 Aug 2017.
- Plüss, B., & De Liddo, A. (2015). Engaging citizens with televised election debates through online interactive replays. In *Proceedings of the ACM international conference on interactive experiences for TV and online video*, (pp. 179–184). ACM.
- Polanyi, L. (1988). A formal model of the structure of discourse. *Journal of Pragmatics*, 12(5), 601–638.
- Prasad, R., Webber, B., & Joshi, A. (2014). Reflections on the penn discourse treebank, comparable corpora, and complementary annotation. *Computational Linguistics*, 40(4), 921–950. [https://doi.org/10.1162/COLI\\_a\\_00204](https://doi.org/10.1162/COLI_a_00204).

- Rahwan, I., Zablith, F., & Reed, C. (2007). Laying the foundations for a world wide argument web. *Artificial Intelligence*, 171, 897–921.
- Reed, C. (2006). Representing dialogic argumentation. *Knowledge-Based Systems*, 19(1), 22–31.
- Reed, C., & Budzyska, K. (2011). How dialogues create arguments. In F. H. van Eemeren, B. Garssen, D. Godden & G. Mitchell (Eds.) *Proceedings of the 7th conference of the international society for the study of argumentation (ISSA), SicSat*
- Reed, C., Mochales Palau, R., Rowe, G., & Moens, MF. (2008a). Language resources for studying argument. In N. Calzolari, K. Choukri, B. Maegaard, J. Mariani, J. Odijk, S. Piperidis, & D. Tapias (Eds.) *Proceedings of the sixth international conference on language resources and evaluation (LREC'08), European Language Resources Association (ELRA), Marrakech, Morocco*
- Reed, C., & Rowe, G. (2004). Araucaria: Software for argument analysis, diagramming and representation. *International Journal on Artificial Intelligence Tools*, 13(4), 961–980.
- Reed, C., Wells, S., Rowe, G., & Devereux, J. (2008b). AIF+: Dialogue in the argument interchange format. In P. Besnard, S. Doutre, & A. Hunter (Eds.) *Proceedings of the 2nd international conference on computational models of argument (COMMA 2008)*, (pp. 311–323). IOS Press.
- Romero-Trillo, J. (2017b). Corpus pragmatics. *Corpus Pragmatics*, 1(1), 1–2. <https://doi.org/10.1007/s41701-017-0005-z>.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4), 696–735.
- Schegloff, E., & Sacks, H. (1974). Opening up closings. In R. Turner (Ed) *Ethnomethodology: Selected readings*, (pp. 223–264). London: Penguin
- Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge: Cambridge University Press.
- Stab, C., & Gurevych, I. (2017). Parsing argumentation structures in persuasive essays. *Computational Linguistics*, 43(3), 619–659. [https://doi.org/10.1162/COLL\\_a\\_00295](https://doi.org/10.1162/COLL_a_00295).
- Stede, M., Afantenos, S., Peldszus, A., Asher, N., & Perret, J. (2016). Parallel discourse annotations on a corpus of short texts. In N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Eds.) *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*.
- Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C., & Lee, L. (2016). Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th international conference on world wide web, international world wide web conferences steering committee, Republic and Canton of Geneva, Switzerland, WWW '16*, (pp. 613–624).
- Teufel, S., Carletta, J., & Moens, MF. (1999). An annotation scheme for discourse-level argumentation in research articles. In *Proceedings of the ninth conference on European chapter of the Association for Computational Linguistics*, (pp. 110–117). Association for Computational Linguistics.
- Teufel, S., Siddharthan, A., & Batchelor, C. (2009). Towards discipline-independent argumentative zoning: Evidence from chemistry and computational linguistics. In *Proceedings of the 2009 conference on empirical methods in natural language processing: Volume 3-Volume 3*, (pp. 1493–1502). Association for Computational Linguistics.
- Vail, AK., & Boyer, KE. (2014). Identifying effective moves in tutoring: On the refinement of dialogue act annotation schemes. In *International conference on intelligent tutoring systems*, (pp. 199–209). Springer
- van Eemeren, F. H., Garssen, B., Krabbe, E. C. W., Snoeck Henkemans, A. F., Verheij, B., & Wagemans, J. H. M. (2014). *Handbook of argumentation theory*. Berlin: Springer.
- van Eemeren, F. H., Houtlosser, P., & Snoeck Henkemans, A. F. (2007). *Argumentative indicators in discourse: A pragma-dialectical study*. Springer: Argumentation Library.
- Visser, J., Duthie, R., Lawrence, J., & Reed, C. (2018a). Intertextual correspondence for integrating corpora. In N. Calzolari, K. Choukri, C. Cieri, T. Declerck, S. Goggi, K. Hasida, H. Isahara, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, S. Piperidis, T. Tokunaga (Eds.) *Proceedings of the eleventh international conference on language resources and evaluation (LREC 2018)*, (pp. 3511–3517). Miyazaki, Japan: European Language Resources Association (ELRA).
- Visser, J., Lawrence, J., Wagemans, J., & Reed, C. (2018b). Revisiting computational models of argument schemes: Classification, annotation, comparison. In S. Modgil, et al. (Eds.) *Proceedings of the 7th international conference on computational models of argument (COMMA 2018)*. Warsaw, Poland: IOS Press



- Walker, M., Anand, P., Abbott, R., Tree, J., Martell, C., & King, J. (2012a). That's your evidence? Classifying stance in online political and social debate. *Decision Support Sciences*, 53(4), 719–729.
- Walker, M. A., Tree, J. E. F., Anand, P., Abbott, R., & King, J. (2012b). A corpus for research on deliberation and debate. In *Proceedings of the 8th edition of the Language Resources and Evaluation Conference (LREC)*, (pp. 812–817).
- Walton, D., & Krabbe, E. (1995). *Commitment in dialogue: Basic concepts of interpersonal reasoning*. New York: State University of New York Press.
- Walton, D., Reed, C., & Macagno, F. (2008). *Argumentation schemes*. Cambridge: Cambridge University Press.
- Wang, Y., Li, Y., & Luo, J. (2016a). Deciphering the 2016 U.S. presidential campaign in the Twitter sphere: A comparison of the Trumpists and Clintonists. In *International AAAI conference on web and social media*. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13064>
- Wang, Y., Luo, J., Niemi, R., Li, Y., & Hu, T. (2016b). Catching fire via “likes”: Inferring topic preferences of Trump followers on Twitter. In *International AAAI conference on web and social media*. <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM16/paper/view/13054>.
- Wei, Z., Liu, Y., & Li, Y. (2016). Is this post persuasive? Ranking argumentative comments in the online forum. In *Proceedings of the 54th annual meeting of the association for computational linguistics*, (vol. 2, pp 195–200). Association for Computational Linguistics.
- Weisser, M. (2014). Speech act annotation. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics: A handbook* (pp. 84–116). Cambridge: Cambridge University Press.
- Weisser, M. (2016). Dart-the dialogue annotation and research tool. *Corpus Linguistics and Linguistic Theory*, 12(2), 355–388.
- Wells, C., Shah, D. V., Pevehouse, J. C., Yang, J., Pelled, A., Boehm, F., et al. (2016). How Trump drove coverage to the nomination: Hybrid media campaigning. *Political Communication*, 33(4), 669–676. <https://doi.org/10.1080/10584609.2016.1224416>.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.